Στατιστική για Ψυχολόγους

Πανεπιστημιακές Σημειώσεις

Βασίλης Βασδέκης

Τμήμα Στατιστικής

Οικονομικό Πανεπιστήμιο Αθηνών

1 Εισαγωγή στις βασικές έννοιες Πιθανοτήτων

Στην παράγραφο αυτή δίνονται συνοπτικά οι βασικότερες έννοιες στη θεωρία πιθανοτήτων καθώς και οι βασικότερες πράξεις. **Πείραμα τύχης** ή **μεταβλητή** είναι κάθε πράξη που οδηγεί σε προσδιοριστέο ποσοτικά αποτέλεσμα. **Ενδεχόμενο** είναι το αποτέλεσμα ενός πειράματος τύχης ή μιας μεταβλητής. Παραδειγμα: η ρίψις ενός νομίσματος είναι πείραμα τύχης με δυο δυνατά ενδεχόμενα, κορώνα ή γράμματα.

Ως **κλασσική πιθανότητα** εμφάνισης ενός ενδεχομένου ορίζεται ο λόγος

$$P(E) = \frac{\text{Αριθμός αποτελεσμάτων που ικανοποιούν το ενδεχόμενο E}}{\text{Ολικό αριθμό αποτελεσμάτων πειράματος}}$$

Ο ορισμός της κλασσικής πιθανότητας που δόθηκε παραπάνω έχει περιορισμούς μια και είναι δύσκολο να καθορίσουμε τον αριθμό δυνατών αποτελεσμάτων σε πειραματικές συνθήκες. Ενας εναλλακτικός ορισμός που δόθηκε, και αυτός όχι χωρίς περιορισμούς, είναι αυτός της στατιστικής πιθανότητας, όπου σε μια σειρά n επαναλήψεων του πειράματος μετράται ο αριθμός που εμαφανίστηκε το ενδεχόμενο E, $n(E)$. Τότε ως στατιστική πιθανότητα του ενδεχομένου ορίζεται να είναι το όριο του λόγου

$$\frac{n(E)}{n}$$

όσο το n, ο αριθμός επαναλήψεων του πειράματος, τείνει στο άπειρο. Σε αυτή τη σειρά μαθημάτων θα χρησιμοποιηθεί ο ορισμός της κλασσικής πιθανότητας.

Παραδείγματα: Ας ορίσουμε τα ενδεχόμενα:
α) E_1: Ρίχνω ένα ζάρι το οποίο φέρνει αποτέλεσμα 6.
β) E_2: Ρίχνω διαδοχικά δυο νομίσματα και το αποτέλεσμα είναι ΚΚ.
γ) E_3: Οπως στο $_2$ με αποτέλεσμα μια φορά Κ.
δ) E_4: Ενα κουτί περιέχει 6 λευκές και 3 μαύρες σφαίρες. Διαλέγω μια σφαίρα στην τύχη. Η σφαίρα βγαίνει λευκή.
Απαντήσεις: α) Τα δυνατά ενδεχόμενα είναι: $\{1, 2, 3, 4, 5, 6\}$. Αρα $P(E_1) = 1/6$.
β) Τα δυνατά ενδεχόμενα είναι: $\{KK, K\Gamma, \Gamma, \Gamma\Gamma\}$. Αρα $(E_2) = 1/4$.
γ) Τα δυνατά ενδεχόμενα είναι όπως στο β). Οπότε $(E_3) = 2/4$.
δ) $P(E_4) = 6/9$.

Σύνθετα ενδεχόμενα μπορούν να εκφρασθούν ως συζεύξεις ή διαζεύξεις απλούστερων ενδεχομένων. Στις περιπτώσεις αυτές ισχύουν τα παρακάτω:

1.1 Προσθετικός Νόμος Πιθανοτήτων

Για δυο ενδεχόμενα E_1 και E_2 ισχύει:

$$P(E_1 \text{ ή } E_2) = P(E_1) + P(E_2) - P(E_1 \text{ και } E_2)$$

Οταν τα δυο ενδεχόμενα είναι **αμοιβαίως αποκλειώμενα** ή **ασυμβίβαστα** τότε δεν είναι δυνατόν να εμφανιστούν μαζί οπότε $P(E_1 \text{ και } E_2) = 0$. Αρα ο προσθετικός νόμος στην περίπτωση αυτή γίνεται:

$$P(E_1 \text{ ή } E_2) = P(E_1) + P(E_2)$$

Παραδείγματα: α) Τραβώ ένα χαρτί από μια τράπουλα: Ζητώ την πιθανότητα Να τραβήξω φιγούρα ή καρρώ.
β) Προηγούμενο παράδειγμα: Ζητώ την πιθανότητα να τραβήξω φιγούρα ή 2.

Απαντήσεις: Εστω E_1: τραβώ φιγούρα, E_2: τραβώ καρρώ και E_3: τραβώ 2. Τότε,
α) $P(E_1) = 12/52$, $P(E_2) = 13/52$, $P(E_1 \text{ και } E_2) = 3/52$. Οπότε,

$$P(E_1 \text{ ή } E_2) = P(E_1) + P(E_2) - P(E_1 \text{ και } E_2) = \frac{12}{52} + \frac{13}{52} - \frac{3}{52} = \frac{22}{52}$$

β) Τα ενδεχόμενα E_1 και E_3 είναι αμοιβαίως αποκλειώμενα. Επίσης, $P(E_3) = 4/52$. Οπότε,

$$P(E_1 \text{ ή } E_3) = P(E_1) + P(E_3) = \frac{16}{52}$$

1.2 Πολλαπλασιαστικός Νόμος Πιθανοτήτων

Δυο ενδεχόμενα καλούνται **ανεξάρτητα** εάν η πραγματοποίηση του ενός δεν έχει καμμιά επίδραση στην πραγματοποίηση του άλλου. Για δυο ενδεχόμενα E_1 και E_2 (χωρίς να προσδιορίζεται αν είναι ανεξάρτητα ή όχι) ισχύει:

$$P(E_1 \text{ και } E_2) = P(E_1/E_2)P(E_2)$$

Η πιθανότητα $P(E_1/E_2)$ είναι η πιθανότητα εμφάνισης του ενδεχομένου E_1 δοθέντος ότι το ενδεχόμενο E_2 έχει ήδη εμφανιστεί. Εαν τα ενδεχόμενα είναι ανεξάρτητα τότε $P(E_1/E_2) = P(E_1)$ οπότε,

$$P(E_1 \text{ και } E_2) = P(E_1)P(E_2)$$

Παραδείγματα: α) Ρίχνω ένα νόμισμα δυο φορές. Ποια η πιθανότητα να φέρω ΚΚ?
β) Μια οικογένεια έχει δυο παιδιά. Δοθέντος ότι η πιθανότητα εμφάνισης αγοριού και κοριτσιού είναι η ίδια ποια η πιθανότητα να έχει η οικογένεια αυτή τουλάχιστον ένα αγόρι δοθέντος ότι ένα παιδί είναι κορίτσι?

Απαντήσεις: α) Εστω E_1: Κ στην πρώτη ρίψη και E_2: Κ στη δεύτερη ρίψη. Τότε τα δυο ενδεχόμενα είναι ανεξάρτητα οπότε,

$$P(E_1 \text{ και } E_2) = P(E_1)P(E_2) = \frac{1}{2}\frac{1}{2} = \frac{1}{4}$$

β) Εστω E_3: τουλάχιστον ένα αγόρι και E_4: ακριβώς ένα κορίτσι. Τότε, $P(E_3 \text{ και } E_4) = 2/4$ και $P(E_4) = 3/4$. Οπότε,

$$P(E_3/E_4) = \frac{P(E_3 \text{ και } E_4)}{P(E_4)} = \frac{2/4}{3/4} = \frac{2}{3}$$

Για ένα πείραμα τύχης ή μια μεταβλητή, το σύνολο των δυνατών ενδεχομένων ονομάζεται **χώρος ενδεχομένων**, ενώ το σύνολο των πιθανοτήτων που αντιστοιχεί σε κάθε ένα από τα ενδεχόμενα ορίζει την **κατανομή πιθανοτήτων** ή την **κατανομή συχνοτήτων** του πειράματος τύχης ή της μεταβλητής. Δυο βασικές παράμετροι μιας κατανομής πιθανοτήτων είναι η **μέση τιμή** και η **διασπορά**. Για μια μεταβλητή X (ή και πείραμα τύχης), η μέση τιμή και η διασπορά συμβολίζονται με $E(X)$ ή μ και $\Delta(X)$ ή σ^2 αντίστοιχα και ορίζονται να είναι

$$E(X) = \mu = \sum xP(X = x) \text{ και } \Delta(X) = \sigma^2 = E(X - \mu)^2 = \sum (x - \mu)^2 P(X = x)$$

Ιδιότητες: Για δυο μεταβλητές X_1 και X_2 και αριθμό (σταθερά) a έχουμε:

- $E(X_1 + X_2) = E(X_1) + E(X_2)$, $E(X_1 - X_2) = E(X_1) - E(X_2)$
- $E(aX_1) = a(X_1)$
- $\Delta(aX_1) = a^2\Delta(X_1)$

Παράδειγμα: Εστω ότι ρίχνουμε δυο ζάρια και ως αποτέλεσμα παίρνουμε το άθροισμα των αποτελεσμάτων των δυο ρίψεων. Τα δυνατά ενδεχόμενα είναι οι αριθμοί από 2 έως 12. Οι πιθανότητες, εύκολα ελέγχεται ότι είναι: $P(2) = 1/36$, $P(3) = 2/36$, $P(4) = 3/36$, $P(5) = 4/36$, $P(6) = 5/36$, $P(7) = 6/36$, $P(8) = 5/36$, $P(9) = 4/36$, $P(10) = 3/36$, $P(11) = 2/36$, $P(12) = 1/36$. Ποιο είναι το μέσο αποτέλεσμα που πρέπει να περιμένουμε όταν εκτελέσουμε το πείραμα μια φορά? Η απάντηση δίνεται από τον αριθμητικό προσδιορισμό της μέσης τιμής. Αν η μεταβλητή συμβολιστεί με X τότε

$$E(X) = 1 \times P(1) + 2 \times P(2) + 3 \times P(3) + \ldots + 12 \times P(12) = \frac{252}{36} = 7$$

Αρα το μέσο αποτέλεσμα που περιμένουμε έιναι το 7. Παράλληλα αυτό τυχαίνει να είναι και το πιο πιθανό χωρίς αυτό να συμβαίνει πάντοτε.

2 Κανονική κατανομή

Στην προηγούμενο τμήμα των σημειώσεων εξετάσαμε μεταβλητές οι οποίες είχαν ως ενδεχόμενα ακέραιες τιμές (**διακριτές μεταβλητές**). Αυτές τις ξεχωρίζουμε από τις **συνεχείς μεταβλητές**, οι οποίες έχουν ως ενδεχόμενα όχι ακέραιους αριθμούς αλλά διαστήματα. Θυμηθείτε ότι όταν η βαθμολογία ενός φοιτητή σε κάποιο μάθημα είναι 8 τότε αυτό το αποτέλεσμα δεν αποτελεί ενδεχόμενο διακριτής κατανομής αλλά συνεχούς μια και αντιπροσωπεύει όλους τους αριθμούς στο διάστημα [7.5,8.5] οι οποίοι κωδικοποιούνται ως 8. Λογικό είναι λοιπόν αν αναμένει κανείς το διάγραμμα πιθανοτήτων να μην είναι σημειακό αλλά να αντιπροσωπεύεται από μια καμπύλη γραμμή.

Οι πιθανότητες ενός ενδεχομένου (διαστήματος) σε μια συνεχή κατανομή βρίσκονται από το εμβαδόν της περιοχής που περικλείεται από την καμπύλη και το διάστημα την πιθανότητα του οποίου ενδιαφερόμαστε να υπολογίσουμε.

Μια από τις βασικές κατανομές που χρησιμοποιείται εκτενώς στη στατιστική και από αυτήν προκύπτουν οι περισσότερες βοηθητικές κατανομές είναι η **κανονική κατανομή** ορισμένη από τον Gauss. Είναι συμμετρική κωδωνοειδής και χαρακτηρίζεται από δυο παραμέτρους. Τη μέση της τιμή (μ) που αποτελεί και το σημείο συμμετρίας της κατανομής και τη διασπορά της (σ^2) που αποτελεί ένα μέτρο του πόσο ανοικτή (διεσπαρμένα σημεία μακριά από το μέσο) ή όχι είναι η κατανομή. Συμβολίζεται με (μ, σ^2). Η κανονική κατανομή $(0, 1)$ καλείται **τυποποιημένη** και αν συμβολίσουμε με μια μεταβλητή που ακολουθεί κανονική κατανομή (μ, σ^2) και με μια μεταβλητή που ακολουθεί τυποποιημένη κανονική τότε υπάρχει η ακόλουθη σχέση που τις συνδέει:

$$Z = \frac{X - \mu}{\sigma}$$

Η πράξη αυτή που γίνεται στη για να προκύψει η λέγεται **τυποποίηση**. Μπορούμε λοιπόν μια οποιαδήποτε κανονική κατανομή να την τυποποιήσουμε. Γι αυτό το λόγο και έχουν κατασκευαστεί πίνακες πιθανοτήτων (για να βρίσκουμε δηλαδή τις πιθανότητες των ενδεχομένων που μας ενδιαφέρουν) μόνο για την τυποποιημένη κανονική κατανομή μια και η πιθανότητα εμφάνισης μιας ατυποποίητης τιμής είναι ίδια με την πιθανότητας εμφάνισης της αντίστοιχης τυποποιημένης.

Παραδείγματα: Εστω ότι η μεταβλητή Z ακολουθεί τυποποιημένη κανονική κατανομή και ότι η X ακολουθεί $N(10,4)$. Ζητούνται:
α) $P(Z<1.3)$, $P(0.8<Z<1.5)$.
β) $P(11.6<X<13)$.

Απαντήσεις: α) $P(Z<1.3)=0.4032$. $P(0.8<Z<1.5)=P(Z<1.5)-P(Z<0.8)=0.4332-0.2881=0.1451$.
β) Οι τιμές 11.6 και 13 είναι τιμές της μη τυποποιημένης κατανομής. Οι τιμές αυτές τυποποιούνται με τη διαδικασία που περιγράφηκε προηγουμένως. Ετσι έχουμε για τις 11.6 και 13 αντίστοιχα:

$$\frac{11.6-10}{2}=0.8 \quad \text{και} \quad \frac{13-10}{2}=1.5$$

Ετσι, $P(11.6<X<13)=P(0.8<Z<1.5)=0.1451$ μια και η X έγινε Z (τυποποιήθηκε).

3 Δειγματοληψία και βασικές στατιστικές συναρτήσεις.

Η Στατιστική βασίζει τα αποτελέσματά της στην παραδοχή ότι ένας ερευνητής έχει στα χέρια του ένα μικρό κομμάτι του πληθυσμού το οποίο έχει την ιδιότητα να θεωρείται μια μικρογραφία του πληθυσμού υπό παρατήρηση. Αυτό πρακτικά σημαίνει ότι τα συμπεράσματα που θα εξαχθούν από την ανάλυση θα είναι επεκτάσιμα σε ολόκληρο τον πληθυσμό. Το μικρό αυτό καμμάτι του πληθυσμού επιλέγεται με διάφορες μεθόδους που σκοπό έχουν να το κάνουν αντιπροσωπευτικότερο του πληθυσμού και κατά συνέπεια ανάλογα και τα αποτελέσματα που θα εξαχθούν. Δυο είναι οι βασικότερες μέθοδοι που χρησιμοποιούνται για την εξαγωγή ενός **δείγματος**: α) η **απλή τυχαία δειγματοληψία** και β) η **δειγματοληψία κατά στρώματα**. Στην απλή τυχαία δειγματοληψία με έναν εντελώς τυχαίο τρόπο επιλέγουμε κάθε φορά και ένα διαφορετικό μέλος του πληθυσμού υπό παρατήρηση, ενώ στη δειγματοληψία κατά στρώματα διαχωρίζουμε τον πληθυσμό σε στρώματα ανάλογα με τα επίπεδα διαφόρων παραγόντων που μας ενδιαφέρουν (π.χ. κοινωνική τάξη, οικονομική κατάσταση κ.λ.π.) και κατόπιν εκτελούμε απλή τυχαία δειγματοληψία σε κάθε ένα στρώμα του πληθυσμού προσέχοντας η αναλογία του πληθυσμού σε κάθε στρώμα να αντιπροσωπεύεται και στο δειγματικό στρώμα.

Μετά τη δειγματοληψία, η στατιστική προχωρά στην εξαγωγή συμπερασμάτων αξιοποιώντας τις πληροφορίες που περιέχονται στο δείγμα με τη βοήθεια αριθμητικών συναρτήσεων των δεδομένων οι οποίες ονομάζονται **στατιστικές συναρτήσεις**. Αν λοιπόν υποθέσουμε ότι έχουμε μια στατιστική συνάρτηση τότε αυτή είναι τυχαία μεταβλητή και άρα έχει μέση τιμή και διασπορά (τις συμβολίζουμε με μ_T και σ_T^2). Τότε η μεταβλητή

$$\frac{T-\mu_T}{\sigma_T}$$

καλείται **τυποποιημένη μεταβλητή** και έχει μέση τιμή 0 και διασπορά 1. Δεν είναι απαραίτητο η στατιστική συνάρτηση να ακολουθεί κανονική κατανομή. Κάθε τυχαία μεταβλητή με αυτή τη μέση τιμή και αυτή τη διασπορά θα καλείται τυποποιημένη μεταβλητή. Η χρησιμότης μιας τέτοιας μεταβλητής έγκειται στο γεγονός ότι οι πιο γνωστές τυποποιημένες κατανομές έχουν πινακοποιηθεί και έτσι κάθε φορά κάποιος τυποποιεί μια οποιαδήποτε τιμή του και κατόπιν ανατρέχει στον πίνακα που τον ενδιαφέρει (Για μια παρουσίαση των πλέον χρησιμοποιούμενων κατανομών στατιστικών κριτηρίων δες στις σελ. 42--47 του βιβλίου.

4 Εκτιμητική

Η Εκτιμητική είναι το πρώτο βήμα για την εξαγωγή συμπερασμάτων στη Στατιστική. Με τη λέξη εκτίμηση εννοούμε την προσπάθεια να ποσοτικοποιήσουμε τις παραμέτρους που *εμείς* θεωρούμε ότι χαρακτηρίζουν την κατανομή του πληθυσμού. Εαν δεχτούμε ότι τα δυνατά ενδεχόμενα του πειράματός μας χαρακτηρίζονται από την κανονική κατανομή τότε πρέπει να εκτιμήσουμε δυο παραμέτρους: τη μέση τιμή και τη διασπορά. Υπάρχουν περιπτώσεις κατανομών όπως η διωνυμική, όπου χρειάζεται να εκτιμηθεί μόνο μια παράμετρος, και αυτή είναι η πιθανότητα επιτυχίας σε ένα πείραμα τα δυνατά ενδεχόμενα είναι δυο, επιτυχία και αποτυχία. Υπάρχουν καλοί θεωρητικοί λόγοι οι οποίοι μας επιβάλλουν την εκτίμηση της μέσης τιμής και της διασποράς μόνο ακόμα και εαν δε γνωρίζουμε τίποτε για τη φύση της κατανομής των ενδεχομένων ενός πειράματος.

Εκτιμήσεις υπάρχουν δυο ειδών: **σημειακές** και **εκτιμήσεις διαστήματος**. Οι σημειακές εκτιμήσεις αποπειρώνται με έναν αριθμό να δείξουν που βρίσκεται η πραγματική τιμή μιας παραμέτρου ενώ οι εκτιμήσεις διαστηματος παρουσιάζουν διαστήματα μέσα στα οποία η πραγματική τιμή μιας παραμέτρου βρίσκεται με μια σχετική βεβαιότητα. Η πιο συνηθισμένη σημειακή εκτίμηση για τη μέση τιμή του πληθυσμού είναι ο δειγματικός μέσος ο οποίος ορίζεται ως παρακάτω:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

όπου $x_1, \ldots, x_n$ είναι το δείγμα και n το μέγεθος του δείγματος. Η αντίστοιχη σημειακή εκτίμηση για τη διασπορά του πληθυσμού ορίζεται παρακάτω:

$$\hat{s}^2 = \sum_{i=1}^{n} \frac{(x_i - \bar{x})^2}{n-1}$$

Στην περίπτωση των εκτιμήσεων διαστήματος το διάστημα το οποίο παράγεται ονομάζεται **διάστημα εμπιστοσύνης** και η βεβαιότητα ή η πιθανότητα να βρίσκεται η αληθινή τιμή της παραμέτρου στο διάστημα αυτό λέγεται **συντελεστής εμπιστοσύνης**. Η τιμή "1-*συντελεστής εμπιστοσύνης*" καλείται **επίπεδο σημαντικότητας**, εκφράζεται σε ποσοστό, συμβολίζεται με "α" και δείχνει το βαθμό για τον οποίο είμαστε αβέβαιοι ως προς το αν βρίσκεται η αληθινή τιμή της παραμέτρου μέσα στο διάστημα εμπιστοσύνης. Η αληθινή ερμηνεία ενός διαστήματος εμπιστοσύνης είναι η παρακάτω: έστω ότι για ένα δείγμα που λαμβάνουμε από τον πληθυσμό, κατασκευάζουμε ένα 95% ΔΕ. Επαναλαμβάνουμε τη διαδικασία παίρνοντας ένα νέο δείγμα από τον πληθυσμό και΄πάλι έτσι ώστε να καταλήξουμε με 1000 (για παράδειγμα) τέτοια ΔΕ. Τότε περιμένουμε το 95% από αυτά να περιέχει την πραγματική τιμή της παραμέτρου του πληθυσμού για την οποία θέλουμε να εξάγουμε συμπέρασμα. Επειδή ωστόσο στην πράξη διαθέτουμε ένα δείγμα, ελπίζουμε ότι το ΔΕ που κατασκευάζουμε είναι ένα από τα 95% ΔΕ που περιέχουν την πραγματική τιμή της παραμέτρου. Στην πραγματικότητα, δεν έχουμε καμία ένδειξη για την πιθανότητα με την οποία το συγκεκριμένο ΔΕ περιέχει την πραγματική τιμή της παραμέτρου. Μια καλή ιδέα είναι ότι τα ΔΕ προτείνουν ένα λογικό εύρος τιμών για την πραγματική τιμή της παραμέτρου, το οποίο μπορεί να εξαχθεί από το δείγμα που διαθέτουμε.

Τα δυο άκρα του διαστήματος λέγονται **όρια εμπιστοσύνης**. Τα ΔΕ μπορούν να είναι **μονοκατάληκτα** ή **δικατάληκτα**. Στα μονοκατάληκτα διαστήματα εμπιστοσύνης η πιθανότητα (αβεβαιότητα) αφαιρείται ολόκληρη από το ένα άκρο του διαστήματος αλλιώς στα δικατάληκτα αφαιρείται μισή-μισή και από τα δυο άκρα. Τα όρια εμπιστοσύνης κατασκευάζονται με τον ακόλουθο τρόπο:

			Κρίσιμη τιμή α%		Τυπικό σφάλμα
Ορια Εμπιστοσύνης =	Τιμή Στατιστικής	±	Κατανομής Στατιστικής		Κατανομής Στατιστικής
	Συνάρτησης		Συνάρτησης		Συνάρτησης

Σύμφωνα λοιπόν με τον παραπάνω κανόνα έχουμε και τα παρακάτω διαστήματα εμπιστοσύνης για τις διάφορες περιπτώσεις στατιστικών συναρτήσεων:

- Δ.Ε. για μέσες τιμές (κανονικός πληθυσμός, σ^2 γνωστό):

$$\bar{x} \pm z_\alpha \frac{\sigma}{\sqrt{n}}$$

  όπου είναι το επίπεδο σημαντικότητος και z ειναι το κρίσιμο σημείο της κανονικής κατανομής στο συγκεκριμένο επίπεδο σημαντικότητας.

- Δ.Ε. για μέσες τιμές (κανονικός πληθυσμός, σ^2 άγνωστο):

$$\bar{x} \pm t_\alpha^{n-1} \frac{\hat{s}}{\sqrt{n}}$$

  όπου είναι το επίπεδο σημαντικότητος και t είναι το κρίσιμο σημείο της t-κατανομής με $n-1$ βαθμούς ελευθερίας.

- Δ.Ε. για αναλογίες (n μεγάλο):

$$p \pm z_\alpha \sqrt{\frac{p(1-p)}{n}}$$

  όπου p είναι η εκτίμηση της αναλογίας και z_α όπως προηγουμένως.

- Δ.Ε. για διαφορές μέσων τιμών (κανονικοί, ομοιογενείς πληθυσμοί):

$$\bar{x}_1 - \bar{x}_2 \pm t_\alpha^{n_1+n_2-2} \sqrt{\hat{s}^2 \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

  όπου x_1 και x_2 είναι οι δειγματικοί μέσοι των δυο δειγμάτων αντίστοιχα και $\hat{s}^2$ η εκτίμηση της κοινής διασποράς των δυο πληθυσμών (δες τμήμα 6). Τα δειγματικά μεγέθη των δυο πληθυσμών είναι n_1 και n_2 αντίστοιχα.

Εαν το Δ.Ε. είναι δικατάληκτο τότε τα κρίσιμα σημεία της t-κατανομής ή της κανονικής κατανομής βρίσκονται από την αντίστοιχη στήλη των δικατάληκτων σημείων στους πίνακες του βιβλίου. Αντίστοιχα, και για τα μονοκατάληκτα.

Παράδειγμα: Εστω ότι δέκα διαφορετικές μετρήσεις της διαμέτρου μιας σφαίρας έδωσαν μέση τιμή $\bar{x} = 4.38$ cm δειγματική τυπική απόκλιση $\hat{s} = 0.06$ cm. Εστω επίσης ότι οι μετρήσεις ακολουθούν κανονική κατανομή. Ενα δικατάληκτο Δ.Ε. σε επίπεδο σημαντικότητας $\alpha = 5\%$ για το πραγματικό μήκος της διαμέτρου της σφαίρας δίδεται από τον τύπο (μια και η τυπική απόκλιση του πληθυσμού εκτιμάται)

$$4.38 \pm t^{9}_{0.05}\frac{0.06}{\sqrt{10}} = 4.38 \pm 0.0429$$

μια και η κρίσιμη τιμή t σε επίπεδο σημαντικότητας 5% και με 9 βαθμούς ελευθερίας είναι 2.26. Το Δ.Ε. παίρνει τη μορφή [4.34,4.42]. Είμαστε λοιπόν βέβαιοι 95% ότι η πραγματική τιμή της αληθινής διαμέτρου αυτής της σφαίρας βρίσκεται μεταξύ 4.34 και 4.42. Με βάση αυτό το δείγμα και υπό την προυπόθεση ότι είναι αντιπροσωπευτικό του πληθυσμού, θα ήταν πολύ λίγο πιθανό (για την ακρίβεια μόνο 5%) να βρίσκεται η αληθινή τιμή της διαμέτρου έξω από το Δ.Ε. Είναι φανερό ότι όσο πιο βέβαιοι θέλουμε να είμαστε για την πραγματική τιμή της διαμέτρου, τόσο πιο πολύ πρέπει να μειώσουμε το επίπεδο σημαντικότητος, άρα και η κρίσιμη τιμή θα μεγαλώσει (όρα στον ανάλογο πίνακα του βιβλίου), άρα θα μεγαλώσει και το Δ.Ε. Αν θέλουμε να είμαστε 100% βέβαιοι για την τιμή της διαμέτρου θα πρέπει να πάρουμε ως Δ.Ε. όλους τους πραγματικούς αριθμούς.

Ενα μονοκατάληκτο Δ.Ε για τη διάμετρο δίδεται αν παρατηρήσουμε ότι η κρίσιμη τιμή για 5% μονοκατάληκτο είναι 1.83, οπότε και ένα μονοκατάληκτο Δ.Ε. προς τα δεξιά θα ήταν

$$4.38 + 1.83 \times \frac{0.06}{\sqrt{10}} = 4.38 + 0.0347 = 4.41$$

δηλ. $[-\infty, 4.41]$.

5 Ελεγχος Στατιστικών Υποθέσεων

Ενα απλό παράδειγμα θα βοηθήσει στην καλύτερη κατανόηση των όρων: *Ενας ερευνητής ερευνά εάν οι συνθήκες διαβίωσης σε ορφανοτροφεία στερούν το παιδί από τη σωστή νοητική τους ανάπτυξη. Λαμβάνει δείγμα από 60 παιδιά τα οποία μεγάλωσαν σε ορφανοτροφεία και βρίσκει το νοητικό μέσο όρο τους* $\bar{x} = 94$. *Γνωρίζει ωστόσο ότι η μέτρηση αυτή ακολουθεί κανονική κατανομή με μέσο* μ *και τυπική απόκλιση* $\sigma = 15$. *Το ερώτημα είναι εαν ουσιαστικά ο νοητικός μέσος όρος των παιδιών που διαβιούν σε ορφανοτροφεία συμπίπτει με τον πληθυσμιακό που είναι ίσος με 100 και η διαφορά αυτή προέκυψε ως αποτέλεσμα της φυσικής διακύμανσης που υφίσταται κάθε δείγμα.*

Το ερώτημα αυτό ειναι ένα κλασσικό παράδειγμα ελέγχου στατιστικών υποθέσεων. Το τυπικό και η ονοματολογία οι οποίες χρησιμοποιούνται είναι οι εξής: Η υπόθεση που προκύπτει από την ερώτηση του ερευνητικού προβλήματος καλείται **Μηδενική Υπόθεση** και εδώ έχει τη μορφή του ερωτήματος "$\mu = 100$" όπου μ ο πραγματικός νοητικός μέσος του πληθυσμού των παιδιών που διαβιούν σε ορφανοτροφεία. Στη Στατιστική η μηδενική υπόθεση έχει πάντοτε και μια ενάντια η οποία ονομάζεται **Εναλλακτική Υπόθεση**. Η εναλλακτική δεν μπορεί να είναι σαφής. Δηλαδή δεν μπορεί να έχει τη μορφή "$\mu = 90$". Αυτό προκύπτει από τη λογική επεξαργασία των στατιστικών υποθέσεων η οποία θα εξηγηθεί παρακάτω. Τρεις είναι οι δυνατές μορφές που μπορεί να έχει η εναλλακτική υπόθεση: $\mu > 100$, $\mu < 100$, $\mu \neq 100$. Στις δυο πρώτες περιπτώσεις όπου ορίζεται η κατεύθυνση η εναλλακτική καλείται **μονοκατάληκτη** αλλιώς είναι **δικατάληκτη**. Η μηδενική υπόθεση είτε απορρίπτεται είτε παραμένει αποδεκτή. Η πορεία της στατιστικής υπόθεσης εξηγείται ως εξής: Η μηδενική θεωρείται ότι ισχύει και εξετάζεται η πιθανότητα των συμπερασμάτων που προκύπτουν από αυτήν την παραδοχή. Εαν η πιθανότητα αυτή είναι μικρή τότε αυτό σημαίνει πως κάτω από την παραδοχή της μηδενικής υπόθεσης τα συμπεράσματα είχαν μικρή πιθανότητα εμφάνισης. Αυτό δε συνηγορεί υπέρ της ισχύος της μηδενικής υπόθεσης οπότε αυτή πρέπει να απορριφθεί.

Το ερώτημα λοιπόν που προκύπτει είναι πόσο μεγάλη θα πρέπει να είναι η διαφορά των υπό σύγκριση τιμών (εδώ του 94 και του 100) ώστε να απορριφθεί ή όχι μια υπόθεση ή αλλιώς, η διαφορά να είναι στατιστικώς **σημαντική** ή **ασήμαντη**. Επειδή δουλεύουμε με πιθανότητες θέτουμε πάντοτε ένα ανώτατο ανεκτό όριο πιθανότητας με την οποία είμαστε αβέβαιοι για το αν η μηδενική υπόθεση, την οποία δεχτήκαμε ως σωστή είναι πράγματι σωστή. Η πιθανότητα αυτή καλείται **επίπεδο σημαντικότητας**, εκφράζεται συνήθως σε ποσοστό συμβολίζεται με $\alpha\%$ και αντιστοιχεί σε ένα σημείο (**κρίσιμη τιμή**) πέρα από το οποίο η διαφορά μεταξύ των υπό σύγκριση τιμών θεωρείται στατιστικώς μεγάλη (λέμε και ότι η διαφορά βρίσκεται στην **περιοχή απόρριψης** αλλιώς βρίσκεται στην **περιοχή αποδοχής**). Γυρνώντας στο παραπάνω παράδειγμα, γνωρίζουμε ο δειγματικός μέσος $\bar{x}$ έχει μέση τιμή τον άγνωστο μέσο του πληθυσμού μ και διακύμανση σ^2/n. Ετσι η στατιστική συνάρτηση $z = \sqrt{n}(\bar{x}-\mu)/\sigma$ ακολουθεί την τυποποιημένη κανονική κατανομή. Για τα δεδομένα του παραδείγματος η τιμή του z είναι -3.09. Εαν θέλαμε να είμαστε αβέβαιοι 5% για το αν η μηδενική υπόθεση της ισότητας $= 100$ είναι λανθασμένη, ενώ έχουμε δεχτεί ότι είναι σωστή, τότε η τιμή για το z θα έπρεπε να είναι μέχρι $z = -1.96$ (σε δικατάληκτο έλεγχο). Η τιμή όμως είναι -3.09 οπότε και κρίνεται ως στατιστικώς σημαντική (ή ότι βρίσκεται στην περιοχή απόρριψης) γιατί εαν η μηδενική υπόθεση ήταν σωστή τότε θα ήμασταν σίγουροι ότι η τιμή του στατιστικού κριτηρίου δεν θα γινόταν λιγότερη από -1.96 κατά 95%. Η τιμή δηλαδή -3.09 είναι κάπως απίθανη γιατί είχε μόνο 5% πιθανότητα να εμφανιστεί γι'αυτό και θεωρούμε ότι έχουμε αρκετές ενδείξεις να απορρίψουμε τη μηδενική υπόθεση. Εαν το δείγμα είχε δώσει $\bar{x} = 97$ τότε $z = -1.55$ οπότε και η διαφορά θα ήταν στατιστικώς ασήμαντη αφού η τιμή αυτή είναι μεγαλύτερη της κρίσιμης τιμής -1.96 που αντιστοιχεί σε επίπεδο σημαντικότητος 5%. Στατιστικώς λοιπόν, σημαντική διαφορά σημαίνει ότι υπάρχει μικρή πιθανότητα η διαφορά που παρατηρείται μεταξύ της τιμής της στατιστικής συνάρτησης και της υποτιθέμενης αληθινής τιμής της παραμέτρου να είναι τυχαία και μάλλον οφείλεται σε κάποιον συστηματικό παράγοντα. Στατιστικώς ασήμαντη είναι η διαφορά για την οποία υπάρχει μεγάλη αντίστοιχη πιθανότητα να είναι τυχαία.

5.1 Είδη σφαλμάτων και διαδικασία ελέγχου

Γίνεται φανερό από τα παραπάνω ότι ακόμα και εαν δουλεύει κανείς με μικρό επίπεδο σημαντικότητας στους ελέγχους πάντα θα υπάρχει η πιθανότητα να απορριφθεί η μηδενική υπόθεση ενώ είναι σωστή ή να παραμείνει αποδεκτή ενώ είναι λανθασμένη. Τα σφάλματα αυτού του είδους καλούνται σφάλματα τύπου I και II αντίστοιχα. Είναι εύκολο να δει κανείς ότι το σφάλμα τύπου I είναι το επίπεδο σημαντικότητας ενώ το σφάλμα τύπου II καλείται και συντελεστής . Το γίνεται μικρότερο όσο το είναι μεγαλύτερο, όσο μεγαλύτερο είναι το μέγεθος του δείγματος ή όσο μικρότερη είναι η διασπορά. **Ισχύς** ενός στατιστικού ελέγχου είναι η πιθανότητα να μη διαπραχθεί σφάλμα τύπου II $(1-\beta)$.

Η διαδικασία του ελέγχου βήμα προς βήμα είναι η παρακάτω:

- Αποφασίζουμε ποιο είναι το στατιστικό κριτήριο που ταιριάζει στον υπό μελέτη έλεγχο
- Υπολογίζουμε την τιμή του από τα δεδομένα
- Βρίσκουμε από τους πίνακες της κατανομής ποια είναι η κρίσιμη τιμή της κατανομής που αντιστοιχεί στο προ-δοθέν επίπεδο σημαντικότητος και στον τύπο του ελέγχου (μονοκατάληκτος ή δικατάληκτος).
- Εαν η τιμή του στατιστικού κριτηρίου είναι μεγαλύτερη (κατ'απόλυτη τιμή μια και μπορούμε να έχουμε αρνητικές τιμές) από την κρίσιμη τιμή, τότε είναι στατιστικώς σημαντική και η υπόθεση απορρίπτεται αλλιώς παραμένει αποδεκτή.

Παρακάτω θα μελετηθούν οι περιπτώσεις που συναντώνται περισσότερο σε τέτοιους ελέγχους.

6 Συγκρίσεις μέσων δείγματος και πληθυσμού

Διακρίνουμε τις παρακάτω περιπτώσεις:
α) Ο πληθυσμός είναι κανονικός και η διασπορά του πληθυσμού είναι γνωστή
β) Ο πληθυσμός είναι κονονικός και η διασπορά του πληθυσμού είναι άγνωστη
γ) Δεν γνωρίζουμε τίποτε για την κατανομή του πληθυσμού και η διασπορά του είναι γνωστή
δ) Δεν γνωρίζουμε τίποτε για την κατανομή του πληθυσμού και η διασπορά του είναι άγνωστη

Η στατιστική συνάρτηση στην περίπτωση α), όπως έχουμε δει και από τα προηγούμενα είναι η

$$z = \frac{\sqrt{n}(\bar{x} - \mu)}{\sigma}$$

Στις περιπτώσεις β) και δ) η άγνωστη διασπορά εκτιμάται από την αντίστοιχη δειγματική τιμή της

$$\hat{s}^2 = \sum_{i=1}^{n} \frac{(x_i - \bar{x})^2}{n-1}$$

Οσον αφορά τώρα την κατανομή της στατιστικής αυτής συνάρτησης σε κάθε μια από τις τέσσερις αυτές περιπτώσεις, όταν η διασπορά εκτιμάται (περιπτώσεις β) και δ)) τότε η κατανομή είναι t^{n-1} (ταυ με $n-1$ βαθμούς ελευθερίας) με τη διαφορά ότι στην περίπτωση δ) το δείγμα πρέπει να είναι μεγάλο (συμβατικά θεωρούμε $n \geq 30$. Στις περιπτώσεις α) και γ) η κατανομή του κριτηρίου είναι τυποποιημένη κανονική με τη διαφορά ότι στην περίπτωση γ) το δείγμα πρέπει να είναι επίσης μεγάλο.

Παραδείγματα: 1) Ζητείται να ελεγχθεί κατά πόσον ένα δείγμα 150 μαθητών είναι ως προς τη νοημοσύνη αντιπροσωπευτικό όλου του πληθυσμού. Ο δειγματικός μέσος υπολογίστηκε σε $\bar{x} = 101.2$ ενώ είναι γνωστό ότι ο πληθυσμιακός μέσος είναι $\mu = 100$, η τυπική απόκλιση $\sigma = 15$ και η κατανομή κανονική. Θέλουμε να ελέγξουμε εάν ο δειγματικός μέσος συμπίπτει στην πραγματικότητα με τον πραγματικό μέσο και η τιμή η οποία υπολογίστηκε δεν είναι παρά απόρροια της φυσικής διακύμανσης του συγκεκριμένου χαρακτηριστικού. Η εναλλακτική υπόθεση είναι δικατάληκτη. Η περίπτωσή μας είναι τυπικά μια ανήκουσα στην περίπτωση α) παραπάνω (διασπορά γνωστή, κανονικός πληθυσμός) και το κριτήριο μας είναι το

$$z = \frac{\sqrt{n}(\bar{x} - \mu)}{\sigma}$$

Η τιμή του z είναι 0.98 και η κρίσιμη τιμή σε επίπεδο σημαντικότητος 5% είναι 1.96 όπως φαίνεται και από τον σχετικό πίνακα της κατανομής. Αρα η τιμή είναι στατιστικώς ασήμαντη και η μηδενική (ο πληθυσμιακός μέσος συμπίπτει με τον δειγματικό) παραμένει αποδεκτή. Από τον τύπο του κεφαλαίου 4 προκύπτουν και τα διαστήματα εμπιστοσύνης τα οποία είναι

$$102.2 \pm 1.96 \times 2.39 = \text{ κάτω άκρο } 98.8, \text{ άνω άκρο } 103.6$$

2) 10 μετρήσεις (από κανονικό πληθυσμό) διαμέτρου μιας σφαίρας έδωσαν $^{-} = 4.38$ cm, $\hat{s} = 0.06$ cm. Εικάζεται ότι η πραγματική της διάμετρος είναι 4.5cm. Ζητείται να ελεγχθεί η ανωτέρω πρότασις.

Γίνεται φανερό ότι η μηδενική μας υπόθεση είναι: μέσος της διαμέτρου του πληθυσμού αυτών των σφαιρών ίσος με τον αριθμό 4.5 έναντι της εναλλακτικής ότι ο μέσος δεν είναι ίσος με 4.5. Η διασπορά του πληθυσμού είναι άγνωστη γι'αυτό και έχει εκτιμηθεί από το $\hat{s}^2$. Το κριτήριο που θα χρησιμοποιηθεί θα είναι το t και η κατανομή η t_9. Υπολογίζουμε την τιμή του κριτηρίου

$$t = \frac{\sqrt{n}(\bar{x} - \mu)}{\hat{s}} = -6.32$$

η οποία σε επίπεδο σημαντικότητας 5% απορρίπτεται μια και το κρίσιμο σημείο που αντιστοιχεί στο συγκεκριμένο επίπεδο σημαντικότητας είναι το -2.26 (βρίσκεται παρά πολυ προς το άκρο της κατανομής του στατιστικού κριτηρίου).

7 Η Περίπτωσις Δυο Δειγμάτων

Γενικά, με τη σύγκριση των μέσων δυο δειγμάτων εννοούμε το στατιστικό έλεγχο που αποφασίζει εαν η διαφορά των δειγματικών μέσων παίρνει τιμή 0 ή όχι. Τα δείγματα ή οι πληθυσμοί οι οποίοι και μελετώνται είναι στην πιο απλή τους μορφή **ανεξάρτητοι**, δηλ. δεν έχουμε λόγους να πιστεύουμε πως τα μέλη των δύο ομάδων σχετίζονται με κάποιο τρόπο μεταξύ τους. Μπορούν ωστόσο, και αυτό είναι μια χρησιμότατη εφαρμογή, να είναι και **εξαρτημένοι** δηλ. μόνο ένα από τα δείγματα να έχει ληφθεί με απλή τυχαία δειγματοληψία ενώ το άλλο να έχει επιλεγεί με τρόπο ώστε κάποιο κάποιο χαρακτηριστικό των μελών του να έχει συνάφεια με το ίδιο χαρακτηριστικό των μελών του πρώτου δείγματος. Για παράδειγμα έστω ότι μας ενδιαφέρει να μελετήσουμε μια μέθοδο γρήγορης απομνημόνευσης. Επιλέγουμε με τυχαίο τρόπο 100 παιδιά από ένα μεγάλο σχολείο κάνουμε ένα test απομνημόνευσης και κατόπιν το επαναλαμβάνουμε στα ίδια παιδιά αφού τα έχουμε εκπαιδεύσει στη μέθοδο αυτή. Τα δύο δείγματα είναι εξαρτημένα διότι οι μετρήσεις παίρνονται πάνω στο ίδιο υποκείμενο (παιδί) και τις δυο φορές με τη διαφορά ότι στη δεύτερη περίπτωση τα αντικείμενα έχουν εκπαιδευτεί σε μια συγκεκριμένη εργασία.

Γίνεται επίσης φανερό ότι τα μεγέθη των δειγμάτων μπορούν να είναι όμοια είτε ανόμοια αλλά και οι διασπορές των δύο πληθυσμών από τους οποίους προέρχονται

τα δείγματα μπορούν να είναι ίσες (ομοιογενείς πληθυσμοί) ή άνισες (ανομοιογενείς πληθυσμοί). Ετσι ξεχωρίζουμε τις παρακάτω περιπτώσεις στους ελέγχους υποθέσεων:
1) Ανεξάρτητα δείγματα και πληθυσμοί ομοιογενείς.
2) Ανεξάρτητα δείγματα και πληθυσμοί ανομοιογενείς.
3) Εξαρτημένα δείγματα (δεν έχει σημασία εαν οι πληθυσμοί είναι ομοιογενείς ή όχι).
Να σημειωθεί ότι δε θα εξετάσουμε την υποπερίπτωση ισοπληθών δειγμάτων μια και αυτή προκύπτει εύκολα από τους τύπους που θα δοθούν αν θέσουμε $n_1 = n_2 = n$. Επίσης βασική υπόθεση αυτής της συζήτησης είναι ότι και οι δυο πληθυσμοί είναι κανονικοί.

Ας θεωρήσουμε την περίπτωση 1) όπου οι πληθυσμοί έχουν κοινή διασπορά. Τότε οι αριθμοί $\hat{s}_1^2$ και $\hat{s}_2^2$ (δειγματικές διασπορές των δυο δειγμάτων αντίστοιχα) αποτελούν εκτιμήσεις της ίδιας ποσότητας. Αρα με συνδυασμό των δυο αυτών εκτιμήσεων μπορούμε να πάρουμε μια καλύτερη εκτίμηση της ποσότητας σ^2 (δες και κεφάλαιο 8 σελ. 73 στο βιβλίο). Ετσι σε περίπτωση ομοιογενών πληθυσμών το κριτήριο t γίνεται:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\hat{s}^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

όπου

$$\hat{s}^2 = \frac{(n_1 - 1)\hat{s}_1^2 + (n_2 - 1)\hat{s}_2^2}{n_1 + n_2 - 2}$$

Το κριτήριο αυτό ακολουθεί, όπως έχουμε δει από το δεύτερο κεφάλαιο, t κατανομή με $n_1 + n_2 - 2$ βαθμούς ελευθερίας.

Παράδειγμα: Θέλουμε να ελέγξουμε το αποτέλεσμα μιας εκπαίδευσης στις δημιουργικές ικανότητες των παιδιών. Λαμβάνουμε δυο διαφορετικά δείγματα από δυο σχολεία για τα οποία έχουμε λόγους να πιστεύουμε ότι δεν υπάρχουν διαφορές μεταξύ των μαθητών, μεγέθους 65 και 45 και βρίσκουμε μέσον όρο βαθμολογίας στους ανεκπαίδευτους $\bar{x}_1 = 7.6$ και στους εκπαιδευμένους $\bar{x}_2 = 8.4$ καθώς και $\hat{s}_1^2 = 2.70$ και $\hat{s}_2^2 = 2.51$. Αποφασίζουμε να παράγουμε μια καλύτερη εκτιμήτρια της κοινής διασποράς του πληθυσμού $\hat{s}^2$:

$$\hat{s}^2 = \frac{(64) \times (2.70) + (44) \times (2.51)}{65 + 45 - 2}$$

Αντικαθιστώντας, το κριτήριο γίνεται

$$t = \frac{7.6 - 8.4}{0.32} = -2.5$$

και αυτό ελέγχεται με τους πίνακες της t κατανομής με $65 + 45 - 2 = 108$ βαθμούς ελευθερίας. Από τους πίνακες γίνεται φανερό ότι η τιμή αυτή είναι στατιστικώς σημαντική (μια και το σημείο που αντιστοιχεί σε επίπεδο σημαντικότητας 5% είναι το 1.98) που σημαίνει ότι η συγκεκριμένη εκπαίδευση έχει θετικές επιπτώσεις στις δημιουργικές ικανότητες των παιδιών.

Στο προηγούμενο παράδειγμα υποθέσαμε κάπως αυθαίρετα ότι οι διασπορές των δυο πληθυσμών είναι ίσες, κάτι το οποίο δεν είναι και η σωστότερη ενέργεια. Ο τυπικός έλεγχος για την ισότητα δυο διασπορών (έλεγχος ομοιογένειας πληθυσμών) είναι ο παρακάτω:
Εστω ότι έχουμε δυο εκτιμήσεις $\hat{s}_1^2$ και $\hat{s}_2^2$ των διασπορών ($^2_1$ και $^2_2$ αντίστοιχα) με αντίστοιχους

βαθμούς ελευθερίας $m-1$ και $n-1$ και θέλουμε να ελέγξουμε ότι αποτελούν εκτιμήσεις της ίδιας ποσότητας. Η μηδενική υπόθεση γίνεται

$$H_0: \qquad \sigma_1^2 = \sigma_2^2$$

και η εναλλακτική

$$H_1: \qquad \sigma_1^2 \neq \sigma_2^2$$

Ο έλεγχος είναι η ποσότητα F οριζόμενη ως εξής:

$$F = \frac{\hat{s}_1^2}{\hat{s}_2^2}$$

η οποία και ακολουθεί κατανομή F με $m-1$ και $n-1$ βαθμούς ελευθερίας. Πρέπει να προσεχθεί στον αριθμητή να μπαίνει πάντα η μεγαλύτερη ποσότητα και βέβαια οι βαθμοί ελευθερίας του αριθμητή να αντιστοιχούν στη συγκεκριμένη ποσότητα. Στο προηγούμενο παράδειγμα ο έλεγχος για την ομοιογένεια των δυο πληθυσμών θα ήταν

$$F = \frac{2.7}{2.51} = 1.08$$

το οποίο είναι στατιστικώς ασήμαντο σε επίπεδο σημαντικότητας 10% μια και η κρίσιμη τιμή από τους πίνακες της κατανομής που αντιστοιχεί σε 64 και 44 βαθμούς ελευθερίας είναι περίπου 1.60.

Ας υποθέσουμε τώρα ότι ο έλεγχος για την ισότητα των δυο διασπορών απέρριψε τη μηδενική υπόθεση (περίπτωση 2)). Τότε το στατιστικό κριτήριο για την ισότητα των μέσων των δυο δειγμάτων είναι το

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\hat{s}_1^2}{n_1} + \frac{\hat{s}_2^2}{n_2}}}$$

Ενώ όμως το κριτήριο της περίπτωσης 1) ακολουθούσε t-κατανομή αυτό θα ακολουθεί (προσεγγιστικά) t-κατανομή αλλά με διορθωμένες τιμές με βάση έναν κανόνα ο οποίος ονομάζεται **διόρθωση Cochran και Cox**: βρίσκουμε τις κρίσιμες τιμές της t-κατανομής για το συγκεκριμένο επίπεδο σημαντικότητας στους βαθμούς ελευθερίας του πρώτου και του δεύτερου δείγματος. Ας τις ονομάσουμε αυτές τις τιμές t_1 και t_2. Τότε η κρίσιμη τιμή με την οποία θα απορρίψουμε ή όχι τον έλεγχο είναι η

$$t' = \frac{\frac{t_1 \hat{s}_1^2}{n_1} + \frac{t_2 \hat{s}_2^2}{n_2}}{\frac{\hat{s}_1^2}{n_1} + \frac{\hat{s}_2^2}{n_2}}$$

Παράδειγμα: Εστω ότι μια μεταβλητή μετριέται σε δυο ομάδες παιδιών μεγέθους $n_1 = 24$ και $n_2 = 36$ αντίστοιχα και βρίσκεται ότι $\bar{x}_1 = 106$, $\bar{x}_2 = 91.8$, $\hat{s}_1^2 = 222$ και $\hat{s}_2^2 = 595.78$. Ζητείται να ελεγχθεί κατά πόσον οι δειγματικοί μέσοι αντιστοιχούν σε ίσους πληθυσμιακούς μέσους.

Ο έλεγχος ομοιογένειας δίνει $F = \frac{595.78}{222} = 2.68$ σε 35 και 23 βαθμούς ελευθερίας. Η κρίσιμη τιμή από τους πίνακες της F-κατανομής είναι 1.93 περίπου και έτσι η διαφορά των δυο διασπορών είναι στατιστικώς σημαντική οπότε και η μηδενική της μη διαφοράς μεταξύ των διασπορών των δυο πληθυσμών απορρίπτεται. Το t-κριτήριο στην περίπτωση αυτή δίδεται από τον παραπάνω τύπο και η τιμή του είναι $t = 2.8$. Οσον αφορά τώρα στην κρίσιμη τιμή με την οποία θα απορρίψουμε ή όχι τη μηδενική υπόθεση της μη διαφοράς μεταξύ των μέσων των δυο πληθυσμών, θα εφαρμόσουμε τη διόρθωση Cochran και Cox. Από τον πίνακα της t-κατανομής βλέπουμε ότι $t_1 = t_{5\%}^{23} = 2.07$ και $t_2 = t_{5\%}^{35} = 2.03$ οπότε η διόρθωση δίνει

$$t' = \frac{19.15 + 33.6}{9.25 + 16.55} = \frac{52.75}{25.8} = 2.04$$

Αρα η τιμή του στατιστικού κριτηρίου είναι στατιστικώς σημαντική και έτσι απορρίπτουμε τη μηδενική υπόθεση της μη διαφοράς μεταξύ των μέσων των δυο πληθυσμών.

Ας περάσουμε τώρα στην περίπτωση των εξαρτημένων ή κατά ζεύγη δειγμάτων. Η τακτική την οποία ακολουθούμε είναι η παρακάτω: Εστω ότι τα δείγματα από κάθε πληθυσμό είναι $x_1, \ldots, x_n$ και $y_1 \ldots, y_n$ αντίστοιχα. Παίρνουμε τις διαφορές μεταξύ των ζευγών και τις συμβολίζουμε με $\delta_1, \ldots, \delta_n$. Δηλαδή, $\delta_1 = x_1 - y_1$ κ.ο.κ. Αυτές οι τιμές αποτελούν και το καινούργιο δείγμα μας. Βρίσκουμε το δειγματικό μέσο των διαφορών αυτών τον οποίο και συμβολίζουμε με $\bar{\delta}$. Βρίσκουμε επίσης κατά το γνωστό τρόπο μια εκτίμηση της διασποράς δηλαδή:

$$\hat{s}_\delta^2 = \frac{\sum_{i=1}^n (\delta_i - \bar{\delta})^2}{n - 1}$$

Η μηδενική τότε υπόθεση μετατρέπεται στην υπόθεση ότι ο μέσος της διαφοράς είναι 0 ή όχι. Αυτό εκτιμάται με το γνωστό κριτήριο t όπως παρακάτω:

$$t = \frac{\bar{\delta}}{\hat{s}_\delta / \sqrt{n}}$$

η τιμή του οποίου και συγκρίνεται κατά τα γνωστά με την κρίσιμη τιμή (σε συγκεκριμένο επίπεδο σημαντικότητας) της t-κατανομής με $n - 1$ βαθμούς ελευθερίας.

Παράδειγμα: Ενας ερευνητής θέλει να ερευνήσει κατά πόσον στα νοητικώς καθυστερημένα παιδιά υπάρχει διαφορά μεταξύ της γλωσσικής και της πρακτικής νοημοσύνης. Ελαβε δείγμα 30 νοητικώς καθυστερημένων παιδιών στα οποία και έδωσε από ένα γλωσσικό και ένα πρακτικό test. Οι μέσες τιμές των γλωσσικών και πρακτικών test αντίστοιχα ήταν $\bar{x}_1 = 63.1$ και $\bar{x}_2 = 67.7$. Η μηδενική υπόθεση εδώ είναι ότι δεν υπάρχει καμιά διαφορά μεταξύ των γλωσσικών και πρακτικών test και η εναλλακτική είναι ότι υπάρχει. Είναι εύκολο να δει κανείς ότι $\bar{\delta} = \bar{x}_1 - \bar{x}_2 = -4.6$. Βρέθηκε επίσης ότι $\hat{s}_\delta^2 = 1.32$ Ετσι το t-κριτήριο πήρε την τιμή -4.00 σε 29 βαθμούς ελευθερίας. Από τους πίνακες της t-κατανομής μπορούμε να δούμε ότι η τιμή είναι στατιστικώς σημαντική (μια και η κρίσιμη τιμή είναι -2.05) δηλαδή η μηδενική υπόθεση της μη διαφοράς μεταξύ γλωσσικής και πρακτικής νοημοσύνης απορρίπτεται.

8 Ανάλυση Διασποράς

Η ανάλυση διασποράς (ANOVA εκ του ANalysis Of VAriance) είναι μια τεχνική η οποία μας επιτρέπει να συγκρίνουμε ταυτόχρονα τις μέσες τιμές πολλών πληθυσμών. Εστω ότι έχουμε k πληθυσμούς και από κάθε ένα παίρνουμε ένα δείγμα (k δείγματα). Ενδιαφερόμαστε για την ισότητα των μέσων των πληθυσμών από τους οποίους προέρχονται τα δείγματα (μηδενική υπόθεση). Η εναλλακτική υπόθεση αναφέρει ότι κάποιοι από τους μέσους πληθυσμιακούς μέσους διαφέρουν (όχι απαραίτητα όλοι). Βασική υπόθεση της τεχνικής είναι οι πληθυσμοί να είναι κανονικοί και ομοιογενείς. Η εφαρμογή της τεχνικής περιλαμβάνει τους παρακάτω υπολογισμούς. Για κάθε δείγμα παίρνουμε το άθροισμα των τιμών του το οποίο και συμβολίζουμε με T_i (δηλ. T_1 είναι το άθροισμα των τιμών του δείγματος 1). Για όλα τα δείγματα μαζί παίρνουμε το ολικό άθροισμα των τιμών τους το οποίο και συμβολίζουμε με . Επίσης παίρνουμε το άθροισμα των τετραγώνων όλων των των τιμών των δειγμάτων μαζί ($\sum_{i,j} y_{ij}^2$). Τότε οι πληροφορίες αυτές μπορούν να ενσωματωθούν στον παρακάτω πίνακα ως εξής: Η ποσότητα $\sum_{i,j} y_{ij}^2 - \frac{T^2}{N}$ όπου $N = n_1 +$

Πληθυσμός 1	Πληθυσμός 2	...	Πληθυσμός k	
y_{11}	y_{21}	...	y_{k1}	
y_{12}	y_{22}	...	y_{k2}	
...	...	...	...	
y_{1n_1}	y_{2n_2}	...	y_{kn_k}	
T_1	T_2	...	T_k	T

$n_2 + \ldots + n_k$, ονομάζεται Ολικό Αθροισμα Τετραγώνων (SS_{TOT}). Μπορεί να δειχθεί τότε ότι το ολικό άθροισμα τετραγώνων χωρίζεται σε δυο ποσότητες οι οποίες ονομάζονται Αθροισμα Τετραγώνων μεταξύ των Ομάδων (ή μεταξύ των πληθυσμών) ($SS_{M.O}$) και Αθροισμα Τετραγώνων εντός των Ομάδων (ή εντός των πληθυσμών) ($SS_{E.O}$). Τα αθροίσματα αυτά των τετραγώνων ορίζονται ως:

$$SS_{M.O} = \frac{T_1^2}{n_1} + \frac{T_2^2}{n_2} + \ldots + \frac{T_k^2}{n_k} - \frac{T^2}{N}$$

και

$$SS_{E.O} = SS_O - SS_{M.O}$$

Κάθε ένα από αυτά τα αθροίσματα τετραγώνων αντιστοιχεί σε ορισμένους βαθμούς ελευθερίας (d.f) οι οποίοι βαθμοί καθώς και οι διαιρέσεις αυτών με τα αθροίσματα τετραγώνων τους (καλούνται Μέσα Τετράγωνα ή MS_{TOT}, $MS_{M.O}$, $MS_{E.O}$ αντίστοιχα) δίδονται από τον παρακάτω πίνακα ο οποίος και καλείται Πίνακας Ανάλυσης Διασποράς: Η τιμή F η οποία εμφανίζεται στην τελευταία στήλη του πίνακα συγκρίνεται με την

Πηγή Διασποράς	SS	d.f	MS	F
Ομάδες	$SS_{M.O}$	k-1	$SS_{M.O}/k-1$	$MS_{M.O}/MS_{E.O}$
Εντός των Ομάδων	$SS_{E.O}$	N-k	$SS_{E.O}/N-k$	
Ολικό	SS_{TOT}	N-1		

κρίσιμη τιμή της $F_{k-1,N-k}$ κατανομής στο επιθυμητό επίπεδο σημαντικότητος. Εαν η τιμή F είναι μεγαλύτερη από την κρίσιμη τιμή τότε συμπεραίνουμε ότι έχουμε λόγους να απορρίψουμε την μηδενική υπόθεση της ισότητας των μέσων και των k πληθυσμών (δειγμάτων). Εαν γίνει αυτό τότε οφείλουμε να αξακριβώσουμε τους πληθυσμούς οι οποίοι και διαφέρουν. Αυτό μπορεί να γίνει με τη μέθοδο των **πολλαπλών συγκρίσεων** του Scheffé. Η μέθοδος αυτή κάνει τις συγκρίσεις ακρίβώς όπως και ο έλεγχος τις ισότητας των μέσων δυο ανεξάρτητων δειγμάτων (δες και σελ. 9) μόνο που στη μέθοδο Scheffé το $\hat{s}^2$ έχει αντικατασταθεί από το $MS_{E.O}$ και η κρίσιμη τιμή δεν δίνεται από την t-κατανομή αλλά από την

$$\sqrt{F_{k-1,N-k}^{\alpha\%}(k-1)}$$

όπου βέβαια % είναι το επιθυμητό επίπεδο σημαντικότητος. Με τον τρόπο αυτό συγκρίνονται όλα τα δυνατά ζεύγη των μέσων και βρίσκονται οι διαφορές μεταξύ τους. Ο έλεγχος αυτός θεωρείται συντηρητικός.

Παράδειγμα: Θέλουμε να εξετάσουμε εαν οι βαθμοί πτυχίου μεταξύ τεσσάρων Πανεπιστημιακών Τμημάτων είναι ίσοι κατά μέσον όρο. Οι πληθυσμοί εδώ προσδιορίζονται από τον παράγοντα Πανεπιστημιακά Τμήματα ο οποίος εδώ έχει τέσσερα επίπεδα. Φιλοσοφική, Νομική, Ιατρική και Φυσικομαθηματική, άρα και τέσσερις πληθυσμούς. Παραθέτουμε παρακάτω τα επιμέρους και ολικά αθροίσματα καθώς και το άθροισμα των τετραγώνων των δεδομένων: $T_1 = 164$ (Φιλοσοφική), $T_2 = 144$ (Νομική), $T_3 = 132$ (Ιατρική), $T_4 = 122$ (Φυσικομαθηματική), $\sum_{ij} y_{ij}^2 = 4008.22$. Δίδεται επίσης ότι το μέγεθος κάθε δείγματος είναι 20 (δηλ. $n_1 = n_2 = n_3 = n_4 = 20$). Εύκολα βλέπει κανείς ότι $= 164 + 144 + 132 + 122 = 562$ και ότι τελικά

$$SS_O = 4008.22 - \frac{562^2}{80} = 60.17$$

$$SS_{M.O} = \frac{164^2}{20} + \frac{144^2}{20} + \frac{132^2}{20} + \frac{122^2}{20} - \frac{562^2}{80} = 48.95$$

$$SS_{E.O} = 60.17 - 48.95 = 11.22$$

Εχουμε λοιπόν τον παρακάτω πίνακα ανάλυσης διασποράς: Η κρίσιμη τιμή σε 5%

Πηγή Διασποράς	SS	d.f	MS	F
Ομάδες	48.95	3	16.32	108.8
Εντός των Ομάδων	11.22	76	0.15	
Ολικό	60.17	79		

επίπεδο σημαντικότητος στην $F_{3,76}$ είναι 2.74. Είναι φανερό ότι υπάρχουν διαφορές μεταξύ των μέσων βαθμών πτυχίου στα τέσσερα αυτά Πανεπιστημιακά Τμήματα έχουμε δηλαδή σοβαρό λόγο να απορρίψουμε τη μηδενική υπόθεση της ισότητας των μέσων. Για να βρούμε τώρα που υπάρχουν οι διαφορές θα εφαρμόσουμε τη μέθοδο Scheffé στη σύγκριση των μέσων των δυο πρώτων πληθυσμών: Από την προηγούμενη συζήτηση έχει γίνει κατανοητό ότι ο έλεγχος για την ισότητα δυο μέσων θα δίνεται από την

$$\frac{\bar{x}_1 - \bar{x}_2}{\sqrt{MS_{E.O}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

όπου $\bar{x}_1$ και $\bar{x}_2$ είναι οι δειγματικοί μέσοι των δυο πληθυσμών και n_1 και n_2 τα αντίστοιχα μεγέθη των δειγμάτων.

$$\bar{x}_1 = \frac{1}{n_1} = 8.2, \quad \bar{x}_2 = \frac{T_2}{n_2} = 7.2$$

Αντικαθιστώντας, λαμβάνουμε τιμή 8.33. Η κρίσιμη τιμή είναι

$$\sqrt{2.74 \times 3} = 2.87$$

Ετσι βλέπουμε ότι ο έλεγχος απορρίπτεται και συμπεραίνουμε ότι πρέπει να υπάρχουν διαφορές μεταξύ των μέσων των πτυχίων των φοιτητών της Φιλοσοφικής και της Νομικής. Συνεχίζοντας, μπορεί κάποιος να κάνει όλες τις δυνατές συγκρίσεις ανά δυο των μέσων και να προσδιορίσει ποιοι μέσοι διαφέρουν και ποιοι όχι.

8.1 Διπλή Ανάλυση Διασποράς

Στην προηγούμενη ενότητα θεωρήσαμε ότι τα επίπεδα ενός μόνο παράγοντα προσδιόριζαν τους πληθυσμούς. Στην παρούσα ενότητα θα θεωρούμε ότι ο συνδυασμός των επιπέδων δυο παραγόντων προσδιορίζει τους πληθυσμούς οπότε και ο έλεγχος της ισότητας των μέσων μεταξύ των πληθυσμών καλείται **Διπλή Ανάλυση Διασποράς**. Η περίπτωση των δυο παραγόντων είναι περισσότερο περίπλοκη από την περίπτωση του ενός παράγοντα. Δεν ελέγχονται μόνο οι μέσοι των επιπέδων των δυο παραγόντων αλλά και το πως ο ένας παράγοντας επηρεάζει τον άλλον. Με τη διπλή ανάλυση διασποράς ελέγχονται τρεις μηδενικές υποθέσεις. Με την πρώτη ελέγχεται εαν οι μέσοι των επιπέδων του πρώτου παράγοντα είναι ίσοι μεταξύ τους. Με τη δεύτερη ελέγχεται εαν οι μέσοι των επιπέδων του δεύτερου παράγοντα είναι ίσοι μεταξύ τους. Με την τρίτη ελέγχεται εαν υπάρχει αλληλεπίδραση μεταξύ των παραγόντων ή αλλιώς αν η διαφορές των μέσων μεταξύ των επιπέδων του πρώτου παράγοντα είναι ίδιες σε όλα τα επίπεδα του δεύτερου παράγοντα.

Η τεχνική υποθέτει ότι τα αποτελέσματα για κάθε συνδυασμό των επιπέδων των δυο παραγόντων είναι διατεταγμένα όπως στον παρακάτω πίνακα: Οπως παρατηρούμε στον

	1ο επίπεδο Β παράγοντα	...	b-οστό επίπεδο Β παράγοντα
1ο επίπεδο Α παράγοντα	y_{111} y_{112} $\vdots$ y_{11n}	$\cdots$ $\cdots$ $\cdots$ $\cdots$	y_{1b1} y_{1b2} $\vdots$ y_{1bn}
$\vdots$	$\vdots$	$\vdots$	$\vdots$
a-οστό επίπεδο Α παράγοντα	y_{a11} y_{a12} $\vdots$ y_{a1n}	$\cdots$ $\cdots$ $\cdots$ $\cdots$	y_{ab1} y_{ab2} $\vdots$ y_{abn}

πίνακα κάθε επίπεδο του Α παράγοντα εμπλέκεται με κάθε ένα επίπεδο του Β. Ετσι εαν στο προηγούμενο παράδειγμα προσθέταμε ως Β παράγοντα το φύλο των φοιτητών θα είχαμε 2 επίπεδα κάθε ένα από τα οποία θα εμπλεκόταν με κάθε επίπεδο του Α παράγοντα δηλ. άνδρες από όλα τα Πανεπιστημιακά Τμήματα και γυναίκες από όλα τα Πανεπιστημιακά Τμήματα. Το επόμενο βήμα είναι να υπολογίσουμε τα μερικά αθροίσματα για κάθε συνδυασμό επιπέδων και κάθε επίπεδο κάθε παράγοντα ξεχωριστά. Τα αθροίσματα αυτά εξηγούνται στον παρακάτω πίνακα: Είναι φανερό από τον παραπάνω

	1ο επίπ. Β παράγ.	...	b-οστό επίπ. Β παράγ.	
1ο επίπ. Α παράγ.	T_{11}	$\cdots$	T_{1b}	T_1
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
a-οστό επίπ. Α παράγ.	T_{a1}	$\cdots$	T_{ab}	T_a
	T_{B1}	$\cdots$	T_{Bb}	T

πίνακα ότι για παράδειγμα, το T_{11} είναι το άθροισμα των n δεδομένων της πρώτης γραμμής και της πρώτης στήλης, το T_{A1} είναι το άθροισμα όλων των δεδομένων της πρώτης γραμμής και το T είναι το άθροισμα όλων των δεδομένων μαζί. Τότε το Ολικό

Αθροισμα Τετραγώνων χωρίζεται ως:

$$SS_{TOT} = SS_{M.O} + SS_{E.O}$$

και το $SS_{M.O}$ ως:

$$SS_{M.O} = SS_{M.ROW} + SS_{.COL} + SS_{.INT}$$

όπου τα τρία τελευταία αθροίσματα τετραγώνων σημαίνουν "Μεταξύ των Γραμμών", "Μεταξύ των Στηλών" και "Μεταξύ των Αλληλεπιδράσεων" αντίστοιχα. Τα αθροίσματα αυτά ορίζονται ως εξής:

$$SS_{TOT} = \sum_{ijk} y_{ijk}^2 - \frac{T^2}{abn}$$

$$SS_{M.ROW} = \sum_{i} \frac{T_{Ai}^2}{bn} - \frac{T^2}{abn}$$

$$SS_{M.COL} = \sum_{j} \frac{T_{Bj}^2}{an} - \frac{T^2}{abn}$$

$$SS_{M.INT} = \sum_{ij} \frac{T_{ij}^2}{n} - \sum \frac{T_{Ai}^2}{bn} - \sum_{j} \frac{T_{Bj}^2}{an} + \frac{T^2}{abn}$$

$$SS_{E.O} = SS_{TOT} - SS_{M.ROW} - SS_{M.COL} - SS_{M.INT}$$

Κατά ακριβή αντιστοιχία με τη Μονή ANOVA ορίζεται ο παρακάτω πίνακας ανάλυσης διασποράς: Η ποσότητα F των "Γραμμών" ελέγχει την ισότητα των μέσων των a επιπέδων

Πηγή Διασποράς	SS	d.f	MS	F
Γραμμές	$SS_{M.ROW}$	$a-1$	$MS_{M.ROW}$	$MS_{M.ROW}/MS_{E.O}$
Στήλες	$SS_{M.COL}$	$b-1$	$MS_{M.COL}$	$MS_{M.COL}/MS_{E.O}$
Αλληλεπίδραση	$SS_{M.INT}$	$(a-1)(b-1)$	$MS_{M.INT}$	$MS_{M.INT}/MS_{E.O}$
Εντός των Ομάδων	$SS_{E.O}$	$ab(n-1)$	$MS_{E.O}$	
Ολικό	SS_{TOT}	$abn-1$		

του παράγοντα που εκφράζεται στις Γραμμές. Η κρίσιμη τιμή είναι η $F^{a\%}_{a-1,ab(n-1)}$, όπου α% είναι το επίπεδο σημαντικότητος. Αντιστοίχως οι ποσότητες F στις "Στήλες" και στις "Αλληλεπιδράσεις" ελέγχουν την ισότητα των μέσων των b επιπέδων του παράγοντα που εκφράζεται στις Στήλες και την ύπαρξη ή όχι αλληλεπιδράσεων μεταξύ του ενός παράγοντα και του άλλου. Για την καλύτερη κατανόηση των αλληλεπιδράσεων δείτε στο βιβλίο σας σελ. 180. Οι κρίσιμες τιμές σε αυτές τις περιπτώσεις δίδονται από τις $F^{a\%}_{b-1,ab(n-1)}$ και $F^{a\%}_{(a-1)(b-1),ab(n-1)}$ αντίστοιχα. Πολλαπλές συγκρίσεις μπορούν να γίνουν και εδώ με τη μέθοδο Scheffé δείχνοντας φυσικά την απαραίτητη προσοχή στους βαθμούς ελευθερίας.

8.2 Ανάλυση Διασποράς Μεταξύ Εξαρτημένων Δειγμάτων

Στην περίπτωση πολλών εξαρτημένων δειγμάτων έχουμε υποκείμενα από κάθε ένα από τα οποία παίρνουμε μετρήσεις σε κάποιες μεταβλητές. Για παράδειγμα μπορούμε να έχουμε μαθητές και να πάρουμε ως δείγμα για κάθε έναν από αυτούς τους βαθμούς του στα τρία τρίμηνα του σχολικού έτους. Δημιουργούμε έτσι τρεις πληθυσμούς οι οποίοι και είναι συσχετισμένοι καθώς αντιπροσωπεύουν τρεις μετρήσεις πάνω στο ίδιο υποκείμενο σε τρεις διαφορετικές χρονικές στιγμές. Στην περίπτωση αυτή και καθώς κάθε υποκείμενο μπορεί να θεωρηθεί και μια ξεχωριστή ομάδα, έχουμε μια ιδιότυπη διπλή ανάλυση διασποράς όπου τα υποκείμενα είναι οι "Γραμμές", και οι βαθμοί των τριών τριμήνων είναι οι "Στήλες". Παρατηρήστε ότι επειδή υπάρχει μια μόνο παρατήρηση σε κάθε συνδυασμό γραμμής και στήλης, τότε $SS_{E.O} = 0$ και έτσι το ρόλο του $SS_{E.O}$ στον πίνακα ανάλυσης της διασποράς τον παίζει το $SS_{M.A}$. Είναι φανερό σε αυτήν την περίπτωση ότι έλεγχος μεταξύ των αλληλεπιδράσεων σε αυτήν την περίπτωση δεν υφίσταται (δεν επαρκούν οι βαθμοί ελευθερίας).

9 Απαραμετρικά Κριτήρια

Υπάρχουν περιπτώσεις όπου ένα πείραμα δεν έχει συγκεκριμένη κλίμακα μέτρησης όπως για παράδειγμα όταν κάποιος επιφορτίζεται με την κρίση του χαρακτήρα ενός ανθρώπου και έτσι αναγκάζεται να βαθμολογήσει με μια κλίμακα από το 1-10 δίνοντας κάθε φορά και έναν ακέραιο αριθμό ως αποτέλεσμα. Είναι φανερό ότι με αυτόν τον τρόπο η διαφορά μεταξύ του 5--6 και του 7--8 δεν είναι πάντοτε ακριβώς προσδιορισμένη Αν έχουμε επίσης πολλούς ανθρώπους που μετρούν το ίδιο χαρακτηριστικό τότε αναμένουμε να υπάρχει και κάποια διαφορά στις προτιμήσεις τους όσον αφορά τον ανθρώπινο χαρακτήρα οπότε αυτό επηρεάζει και τις αποφάσεις τους. Σε αυτές τις περιπτώσεις χρησιμοποιούμε έναν **απαραμετρικό έλεγχο**. Τέτοιοι υπάρχουν πολλοί, εδώ θα συναντήσουμε τέσσερις:

9.1 Κριτήριο Mann-Whitney ή U-κριτήριο

Χρησιμοποιείται για τη σύγκριση των μέσων δυο ανεξάρτητων δειγμάτων των οποίων τα μεγέθη είναι n_1 και n_2 αντίστοιχα. Η μηδενική υπόθεση είναι ότι οι αντίστοιχοι πληθυσμιακοί μέσοι είναι ίσοι. Το κριτήριο κατασκευάζει την ποσότητα

$$z = \frac{U - u}{s_u}$$

όπου

$$u = \frac{n_1 n_2}{2} \quad s_u = \sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}}$$

και το U κατασκευάζεται ως εξής: Ενώνουμε τα δυο δείγματα και διατάσσουμε τις τιμές δίνοντας στον μικρότερο αριθμό την τιμή "1" κ.ο.κ. Ονομάζουμε R_1 και R_2 το άθροισμα των τακτικών τιμών του 1ου και του 2ου δείγματος αντίστοιχα. Τότε ο υπολογισμός του U ακολουθεί τον παρακάτω κανόνα:

$$U = n_1 n_2 + \frac{n_1(n_1+1)}{2} - R_1 \quad R_1 < R_2$$
$$U = n_1 n_2 + \frac{n_2(n_2+1)}{2} - R_2 \quad R_2 < R_1$$

Εαν τα n_1, n_2 είναι μεγαλύτερα από το 5, τότε η ποσότητα z ακολουθεί προσεγγιστικά κανονική κατανομή και έτσι ψάχνει κανείς σε αυτόν τον πίνακα για την ανεύρεση της κρίσιμης τιμής (στο επιθυμητό επίπεδο σημαντικότητος).

9.2 Κριτήριο Wilcoxon

Οταν τα δείγματα (μεγέθους n) είναι εξαρτημένα τότε το κατάλληλο κριτήριο για τον έλεγχο της ισότητας των μέσων των δυο πληθυσμών είναι το κριτήριο Wilcoxon. Σύμφωνα με αυτό, κατασκευάζεται η ποσότητα

$$z = \frac{T - \mu}{s_T}$$

όπου

$$\mu = \frac{n(n+1)}{4}, \quad s_T = \sqrt{\frac{n(n+1)(2n+1)}{24}}$$

Οσον αφορά το , παίρνουμε τις διαφορές όπως και στο αντίστοιχο παραμετρικό κριτήριο και διατάσσουμε τις **απόλυτες** τιμές του δίνοντας 1 στη μικρότερη. Η ποσότητα είναι το άθροισμα των τακτικών τιμών των διαφορών των οποίων το πρόσημο έχει τη μικρότερη συχνότητα. Για τιμές του n άνω του 5, το z ακολουθεί προσεγγιστικά κανονική κατανομή.

Σε περίπτωση που υπάρχουν μηδενικές διαφορές, αυτές δε λαμβάνουν μέρος στην ανάλυση.

Παράδειγμα: Θέλουμε να δούμε εαν οι πατέρες ή οι μητέρες γνωρίζουν καλύτερα τις ικανότητες των παιδιών τους. Πήραμε δείγμα 10 οικογενειών που έχουν παιδιά και δώσαμε ένα ερωτηματολόγιο σε κάθε ένα. Ζητήθηκε κατόπιν από τους γονείς των παιδιών να προβλέψουν το σωστό αριθμό απαντήσεων που έδωσε το παιδί τους. Τα αποτελέσματα μαζί με τις διαφορές Δ τις απόλυτες τιμές και τις τακτικές τιμές δίδονται στον πίνακα παρακάτω: Το πρόσημο με τη μικρότερη συχνότητα είναι το "-" οπότε και

Υποκείμενα	Μητέρες	Πατέρες	Δ	$\|\Delta\|$	Τακτικές Τιμές
1	27	13	14	14	9
2	32	13	19	19	10
3	19	27	-8	8	5
4	6	10	-4	4	2
5	27	17	10	10	7
6	31	22	9	9	6
7	16	11	5	7	3.5
8	15	17	-2	2	1
9	18	23	-5	5	3.5
10	23	10	13	13	8

το προκύπτει ως $= 5 + 2 + 1 + 3.5 = 11.5$, $= \frac{10 \times 11}{4} = 27.5$ και $s_T = 9.81$. Τότε

$$z = \frac{11.5 - 27.5}{9.81} = -1.63$$

Το κρίσιμο σημείο της κανονικής κατανομής σε 5% επίπεδο σημαντικότητος είναι το -1.96 οπότε και συνάγομε ότι δεν υπάρχει διαφορά μεταξύ των προβλέψεων των δυο γονέων για τις δυνατότητες των παιδιών τους.

9.3 Κριτήριο Kruskall-Wallis

Στην περίπτωση πολλών (έστω k) ανεξάρτητων δειγμάτων (μεγέθους $n_1, \ldots, n_k$) το απαραμετρικό κριτήριο της Ανάλυσης Διασποράς είναι το κριτήριο Kruskall-Wallis. Σύμφωνα με αυτό, ενώνουμε τους k πληθυσμούς και διατάσσουμε το ενιαίο δείγμα δίνοντας "1" στη μικρότερη τιμή. Παίρνουμε τα αθροίσματα των τακτικών τιμών του κάθε δείγματος και τα ονομάζουμε $R_1, R_2, \ldots, R_k$. Αν συμβολίσουμε $n = n_1 + n_2 + \ldots + n_k$ τότε κατασκευάζουμε την ποσότητα ως:

$$H = \left(\frac{12}{n(n+1)} \sum_{i=1}^{k} \frac{R_i^2}{n_i} \right) - 3(n+1)$$

Για τιμές των $n_1, \ldots, n_k$ μεγαλύτερες του 5, το ακολουθεί (χ^2_{k-1}) ($k-1$ βαθμοί ελευθερίας).

Παράδειγμα: Στον παρακάτω πίνακα μπορούμε να δούμε το βαθμό ικανοποίησης 26 υποκειμένων από τη ζωή (με άριστα το 100): Τα 4 δείγματα ενώθηκαν, διατάχθηκαν και

Αγαμοι	Εγγαμοι με παιδιά	Εγγαμοι χωρίς παιδιά	Διαζευγμένοι
53	68	31	15
24	72	68	30
33	55	34	42
45	44	39	16
49	63	48	26
58	56	46	
29	74		
	71		

οι τακτικές τιμές κάθε δείγματος δίνονται στον παρακάτω πίνακα: Πολύ εύκολα βλέπει

Αγαμοι	Εγγαμοι με παιδιά	Εγγαμοι χωρίς παιδιά	Διαζευγμένοι
3	12	7	1
5	18	9	2
8	19	10	4
13	21	14	6
16	22.5	15	11
17	24	22.5	
20	25		
	26		

κανείς ότι $n_1 + \ldots + n_4 = 26$, $R_1 = 82$, $R_2 = 167.5$, $R_3 = 77.5$, $R_4 = 24$. Τότε

$$H = \frac{12}{26 \times 27} \left(\frac{82^2}{7} + \frac{167.5^2}{8} + \frac{77.5^2}{6} + \frac{24^2}{5} \right) - 3 \times (26+1) = 14.45$$

Η κρίσιμη τιμή βρίσκεται από τον πίνακα της χ^2_3 σε επίπεδο σημαντικότητος 5% και είναι 9.35. Ετσι η τιμή της είναι σημαντική και διαπιστώνουμε ότι υπάρχουν διαφορές μεταξύ των τεσσάρων ομάδων. Φαίνεται ότι οι έγγαμοι με παιδιά είναι περισσότερο ευχαριστημένοι από τη ζωή.

9.4 Κριτήριο Friedman

Οταν τα δείγματα είναι εξαρτημένα τότε χρησιμοποιούμε το κριτήριο Friedman. Για κάθε ένα από τα n υποκείμενα χωριστά μετατρέπουμε σε τακτικές τιμές δίνοντας "1" στη μικρότερη. Στη συνέχεια παίρνουμε το άθροισμα των τακτικών τιμών κάθε ομάδας $R_1, R_2 \ldots, R_k$ και το κριτήριο είναι

$$\chi^2 = \frac{12}{nk(k+1)} \sum_{i=1}^{k} R_i^2 - 3n(k+1)$$

Το χ^2 ακολουθεί χ^2_{k-1} και εκεί πρέπει να αναζητηθεί η κρίσιμη τιμή για τον έλεγχο της ισότητας των μέσων των k (εξαρτημένων) δειγμάτων.

10 Ελεγχοι Διαφοράς Ποσοστιαίων Αναλογιών

10.1 Ανεξάρτητα Δείγματα

Εστω δυο δείγματα των οποίων τα αποτελέσματα είναι δίτιμα. Οι τιμές αυτές αναπαριστούν επιτυχία ή αποτυχία σε κάποιο πείραμα. Εστω $p_1,\ p_2$ τα ποσοστά επιτυχίας στα δυο δείγματα. Ενδιαφερόμαστε για τον έλεγχο της ισότητας των ποσοστών αυτών στον πληθυσμό. Η βασική υπόθεση είναι (κάτω από τη μηδενική υπόθεση) ότι τα δυο ποσοστά είναι ίσα οπότε όπως και στην περίπτωση της σύγκρισης δυο μέσων δυο πληθυσμών στον παρονομαστή οφείλουμε να παράγουμε μια κοινή εκτίμηση του κοινού p. Το κατάλληλο στατιστικό κριτήριο είναι

$$z = \frac{p_1 - p_2}{\sqrt{p(1-p)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

όπου

$$p = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2}$$

Το κριτήριο αυτό για σχετικώς μεγάλες τιμές των $n_1,\ n_2$, ακολουθεί κανονική κατανομή.

Παράδειγμα: Εικάζεται ότι στο δημοτικό σχολείο το ποσοστό των αγοριών και των κοριτσιών με θετική στάση στα ελεύθερα αναγνώσματα είναι ίδια. Προς τούτο χρησιμοποιήθηκαν δυο δείγματα αγοριών και κοριτσιών αντίστοιχα μεγεθών $n_1 = 80,\ n_2 = 70$. Τα ποσοστά θετικής στάσης ήταν για το δείγμα των αγοριών $p_1 = 0.622$ και για το δείγμα των κοριτσιών $p_2 = 0.7$. Το ερώτημα της σημαντικής ή όχι διαφοράς απαντάται χρησιμοποιώντας το προηγούμενο στατιστικό κριτήριο. Οι κοινές τιμές p και z βρίσκονται να είναι ίσες με

$$p = \frac{80 \times 0.622 + 70 \times 0.7}{150} = 0.658, \quad z = -1.01$$

η οποία τιμή κρίνεται ως μη σημαντική σε επίπεδο σημαντικότητος 5% μια και η κρίσιμη τιμή της κανονικής κατανομής είναι -1.96.

10.2 Εξαρτημένα Δείγματα

Ενα αρχικό δείγμα από n υποκείμενα υπόκεινται σε κάποιο πείραμα του οποίου το αποτέλεσμα είναι δίτιμο (επιτυχία ή αποτυχία). Αμέσως μετά όλα τα υποκείμενα δέχονται κάποια εκπαίδευση και υπόκεινται πάλι στο ίδιο πείραμα. Από τα $n_{1.}$ υποκείμενα που εμφάνισαν επιτυχία την πρώτη φορά, n_{11} ξαναεμφάνισαν επιτυχία και n_{12} δεν εμφάνισαν. Από τα $n_{.2}$ υποκείμενα που εμφάνισαν αποτυχία την πρώτη φορά, n_{21} εμφάνισαν επιτυχία και n_{22} εμφάνισαν αποτυχία. Είναι προφανές πως $n = n_{1.} + n_{.2}$. Το ερώτημα σε αυτού του είδους τα προβλήματα είναι εαν το ποσοστό της επιτυχίας πριν και μετά την εκπαίδευση είναι το ίδιο ή όχι. Είναι προφανές ότι οι δυο πληθυσμοί σε αυτή την περίπτωση είναι ουσιαστικά ένας (επαναλαμβανόμενη μέτρηση) άρα είναι και συσχετισμένοι. Το στατιστικό κριτήριο που χρησιμοποιείται σε αυτήν την περίπτωση είναι ο έλεγχος Mc Nemar όπως φαίνεται παρακάτω:

$$z = \frac{n_{12} - n_{21}}{(n_{12} + n_{21})^{1/2}}$$

Και στην περίπτωση αυτή το κριτήριο ακολουθεί κανονική κατανομή για σχετικώς μεγάλες τιμές του n.

10.3 Ελεγχος Καλής Προσαρμογής ^{2} (Goodness-of-fit test)

Για να ελέγξουμε εαν ένα σύνολο συχνοτήτων προέρχεται από μια συγκεκριμένη κατανομή ή όχι, χρησιμοποιούμε τον έλεγχο καλής προσαρμογής χ^2. Για παράδειγμα, θέλουμε να διαπιστώσουμε εαν τα παιδιά μικρής ηλικίας έχουν μια συγκεκριμένη προτίμηση προς κάποια χρώματα. Για το σκοπό αυτό συγκροτήθηκε ένα δείγμα 300 παιδιών και καταγράφηκαν οι προτιμήσεις τους σε πέντε χρώματα. Εαν δεν υπήρχε καμία συγκεκριμένη προτίμηση σε κάποιο ή κάποια χρώματα τότε θα ανέμενε κανείς τα παιδιά να δείξουν ίδια προτίμηση για κάθε χρώμα δηλ. κάθε χρώμα να προτιμηθεί από 60 παιδιά. Τα αποτελέσματα που πήραμε καθώς και οι αναμενόμενες (**Θ**εωρητικές) τιμές σε περίπτωση μη συγκεκριμένης προτίμησης δίδονται στον παρακάτω πίνακα: Ο έλεγχος

Συχνότητες	Χρώματα				
	Καφέ	Κίτρινο	Κόκκινο	Μπλε	Πράσινο
Πραγματικές (Π)	53	44	72	63	68
Αναμενόμενες (Θ)	60	60	60	60	60

της καλής προσαρμογής ως προς τις θεωρητικές τιμές δίδεται από το στατιστικό κριτήριο χ^2 το οποίο ορίζεται

$$\chi^2 = \sum \frac{(\Pi - \Theta)^2}{\Theta}$$

και το οποίο ακολουθεί χ^2_{k-1} όπου k είναι ο αριθμός των ομάδων των οποίων οι συχνότητες εξετάζονται. Στην περίπτωση όπου $k = 2$ τότε χρησιμοποιείται η διόρθωση **Yates** (διόρθωση συνέχειας) η οποία δίδεται παρακάτω:

$$\chi^2 = \sum \frac{\left(|\Pi - \Theta| - \frac{1}{2}\right)^2}{\Theta}$$

Στο παράδειγμα που δώσαμε το κριτήριο υπολογίζεται ως εξής:

$$\chi^2 = \frac{(53-60)^2}{60} + \frac{(44-60)^2}{60} + \ldots + \frac{(68-60)^2}{60} = \frac{522}{60} = 8.7$$

Οι βαθμοί ελευθερίας είναι $k - 1 = 5 - 1 = 4$ και η κρίσιμη τιμή βρίσκεται από τους πίνακες της $^2_4$ κατανομής σε 5% επίπεδο σημαντικότητος η οποία και είναι 9.49. Η τιμή 8.7 κρίνεται μη σημαντική και δείχνει ότι δεν υπάρχει διαφορά ανάμεσα στις προτιμήσεις των παιδιών στα διάφορα χρώματα (στα συγκεκριμένα αυτά χρώματα).

10.4 Πίνακες Συνάφειας

Οι πίνακες συνάφειας χρησιμοποιούνται στην καταχώρηση δεδομένων τα οποία προέρχονται από το συνδυασμό των επιπέδων μιας διμεταβλητής. Ως παράδειγμα αναφέρεται η συσχέτιση ηλικίας παιδιού και προτιμώμενου χρώματος. Δείγμα 300 παιδιών ηλικίας από 3 έως 5 χρόνια εξετάστηκε για το προτιμώμενο χρώμα. Συγκεκριμένα εξετάστηκαν 98 παιδιά ηλικίας 3 ετών, 100 παιδιά ηλικίας 4 ετών και 102 παιδιά ηλικίας 5 ετών. Τα αποτελέσματα φαίνονται στον παρακάτω πίνακα. Τα στοιχεία της τελευταίας γραμμής

Ηλικία	Χρώματα					
	Καφέ	Κίτρινο	Κόκκινο	Μπλε	Πράσινο	
3ο	17	14	24	21	22	98
4ο	16	15	26	20	23	100
5ο	20	15	22	22	23	102
	53	44	72	63	68	300

είναι οι **συχνότητες περιθώριας στήλης** ενώ τα αντίστοιχα της τελευταίας στήλης είναι οι **συχνότητες περιθώριας γραμμής**. Το ερώτημα που εξετάζεται εδώ είναι αν υπάρχει συνάφεια μεταξύ της ηλικίας και του προτιμώμενου χρώματος. Αν δηλαδή, καθώς μεγαλώνει η ηλικία υπάρχει μια σαφής προτίμηση προς κάποιο ιδιαίτερο χρώμα. Στο προηγούμενο μέρος του κεφαλαίου εξετάστηκε το ερώτημα εαν υπάρχει διαφορά μεταξύ των χρωμάτων ανεξαρτήτου ηλικίας. Τώρα θεωρώντας και την ηλικία προσπαθούμε να ανακαλύψουμε κάποια συσχέτιση (ή και συνάφεια) μεταξύ ηλικίας και χρώματος. Το στατιστικό κριτήριο που χρησιμοποιείται είναι το χ^2 το οποίο ορίζεται όπως και προηγουμένως ως:

$$\chi^2 = \sum \frac{(\Pi - \Theta)^2}{\Theta}$$

όπου οι **Θ**εωρητικές (ή και αναμενόμενες) τιμές υπολογίζονται ως εξής:

$$\Theta = \frac{\text{Συχνότητα περιθώριας Στήλης} \times \text{Συχνότητα περιθώριας Γραμμής}}{N}$$

όπου είναι το σύνολο των μελών του δείγματος. Η θεωρητική λοιπόν συχνότητα της ηλικίας του 3ου έτους και χρώματος καφέ είναι $\frac{53\times98}{300} = 17.3$. Με ανάλογο τρόπο υπολογίζονται και οι υπόλοιπες θεωρητικές τιμές για κάθε συνδυασμό των επιπέδων της διμεταβλητής. Το χ^2 υπολογίζεται να είναι

$$\chi^2 = \frac{(17-17.3)^2}{17.3} + \ldots + \frac{(23-23.1)^2}{23.1} = 0.93$$

Το κριτήριο αυτό ακολουθεί $\chi^2_{(\text{αρ.γραμμών-1})\times(\text{αρ.στηλών-1})}$. Εδώ η κατανομή γίνεται χ^2_8. Από τους πίνακες γίνεται φανερό ότι η κρίσιμη τιμή σε 5% επίπεδο σημαντικότητος είναι 15.51, οπότε και η τιμή του χ^2 κρίνεται μη σημαντική. Δεν προκύπτουν λοιπόν ενδείξεις για συσχέτιση της ηλικίας και συγκεκριμένη προτίμηση χρώματος.

Σημειώνουμε ότι σε περίπτωση που ο αριθμός των στηλών και ο αριθμός των γραμμών είναι ίσος με 2 τότε βέβαια οι βαθμοί ελευθερίας του χ^2 είναι 1 και τότε συνιστάται η χρησιμοποίηση της διόρθωσης κατά Yates χωρίς αυτό να αποτελεί και κανόνα.